Assessing Capacity Improvements by Relaying in Cellular Networks

Hans-Florian Geerdes

Konrad-Zuse-Zentrum für Informationstechnik, Berlin, Germany

Abstract. Relaying – allowing multiple wireless hops – is a protocol extension for cellular networks conceived to improve data throughput. Its benefits have only been quantified for small example networks. For assessing its general potential, we define a complex resource allocation/scheduling problem. Several mathematical models are presented for this problem; while a time-expanded MIP approach turns out intractable, a sophisticated column generation scheme leads to good computational results. We thereby show that for selected cases relaying can increase data throughput by 30% on the average.

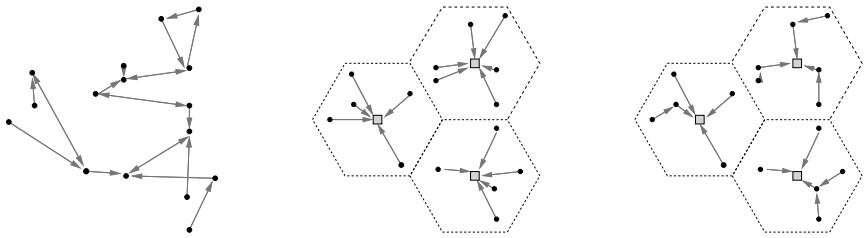
1 Introduction

The amount of data that can be sent per time in wireless computer networks (wireless LANs, WLANs) is limited by bandwidth restrictions and by interference. Many researchers strive to use the available resources more efficiently in order to increase data throughput. One technique conceived to do so is *relaying*.

1.1 Relaying in Cellular Networks

Wireless telecommunication networks fall into two categories: *ad hoc* networks (Fig. 1(a)) and *cellular* networks (Fig. 1(b)). In ad hoc networks, the mobile nodes organize their usage of the radio network themselves, whereas in cellular networks this task is carried out by an infrastructure of fixed *base stations*. In contrast to ad hoc networks, this implies that mobile nodes may never exchange data directly, they are only allowed to transmit data to and receive data from their respective base station.

Relaying means to allow multiple wireless hops for data to reach its destination. While this is quite common in most ad hoc networks, it is normally not used in cellular ones. To use relaying in cellular computer networks can be beneficial because transmission speed can be adapted – several fast hops might be better on the overall than one slow one. Within this work, however, we still rely on the the main feature of cellular network organization, namely organization in space and in time by the base station infrastructure.



(a) Ad hoc network (b) Cellular network (c) Relaying in cellular networks

Fig. 1. Mixing communication paradigms

1.2 The Potential of Relaying

For special network arrangements, it has been shown that using relaying can increase the data throughput in particular cases by up to 60% [3]. The method presented there to analyze the achievable data throughput gain is to find best-possible transmission schedules a) with and b) without relaying and compare the resulting data throughput. However, the *average* achievable additional data throughput by allowing relaying as well as a method to solve the scheduling and resource allocation problem for general networks are yet unknown. This work develops linear and mixed integer programming models for the problem as well as adapted algorithms and uses these tools to assess the *average* potential of relaying in *arbitrary* networks. Some work on scheduling and routing in ad hoc networks uses similar models of the problem, see e.g. [1,6,5], a literature survey can be found in [2].

2 Formulating the Optimization Problem

We now outline the scheduling and resource allocation problem we solve to assess the potential of relaying in cellular networks. The details of our model and its computational parameters are based on HiperLAN/2 radio technology [4]. The results, however, apply essentially to any other modern WLAN technology.

2.1 Transmission Schedules and Data Throughput

Our objective is to maximize data throughput in the *uplink*, that is, we want to transmit as much data to the base station as possible in a fixed period of time of T unit time slots. (We assume that each mobile has an unlimited amount of data to send to its base station.) Data throughput is assigned to the mobiles by specifying a global *transmission schedule* for the network.

A transmission schedule consists of a set of *transmission commands*, that specify all parameters for transmitting a single, atomic *data unit*. For each transmission we have to choose the sender and receiver, the transmission power, and the transmission speed. The last parameter can actually not be chosen freely, but there is a fixed set M of physical modes (six in HiperLAN/2) to pick from. The number of time slots needed to transmit one data unit varies for the different physical modes, we will denote it by $l_m \in \mathbb{N}$ for a mode $m \in M$. The transmission power has to be picked from a (hardware dependent) interval $[\mathrm{P}_{\min}, \mathrm{P}_{\max}]$.

An important limitation to the feasibility of transmission schedules is posed by interference. The amount of interference is normally measured relative to the strength of the desired signal (Signal-to-Interference-and-Noise Ratio, SINR). We simplify the relation between SINR and transmission success and assume that a transmission is successful if and only if the SINR stays above a certain threshold ξ, that is,

$$\frac{\text{Received Signal}}{\sum \text{Interfering Signals} + \nu} \geq \xi_m \,. \tag{SINR}$$

The term ν here stands for an omnipresent amount of thermal noise. The threshold ξ_m varies, it is higher for faster physical modes – higher transmission speed comes at the cost of increased error liability.

2.2 Fairness

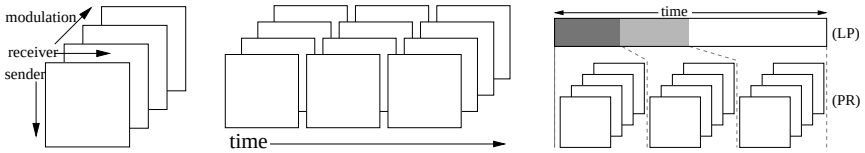
If we only strive to maximize the amount of data that is transmitted to the base station, the resulting schedules will be quite unbalanced: stations that are close to the base station will get all the data throughput (and we would never observe relaying). As this behavior is unwanted, we need to add fairness conditions to our problem that ensure that each terminal gets a share of the overall throughput. The actual modeling of fairness is crucial; different fairness models will lead to quite different results. We have implemented two different notions that form the two extremes of the spectrum in which fairness can be conceived:

Time Share Fairness. No terminal should get more throughput than it would get if the available time was split equally between all mobiles and transmission conditions were optimal (that is, it could use the fastest physical mode). On the other hand, no user should get less throughput than it would get under fair time share if it had to use the slowest physical mode.

Total User Fairness. All stations get the same throughput (and this common throughput is to be maximized). This corresponds roughly to maximizing the minimum throughput that any station gets.

3 Mathematical Models

We now describe mathematical models for the relaying problem. We start with a mixed integer model of feasible transmission conditions at a single instance in time; we then develop two different complete models that rely on this basic model and incorporate the time component. The structure and relation between our models is outlined in Fig. 2.



(a) Transmission Feasibility (b) Time–Expanded MIP (c) Column Model

Fig. 2. Structure of Models

3.1 Feasible Transmission Conditions: a Basic Mixed Integer Model

All possible transmissions are denoted by binary variables $x \in \{0,1\}$, and the associated transmission power is specified in related variables p. If $x = 1$, this means that the referring transmission takes place, otherwise it is not scheduled. We denote the set of all stations by S and the set of available physical modes by M. The index set of all possible transmissions – sketched in Fig. 2(a) – is

$$I := M \times S \times S\,.$$

We have as variables

$$\begin{aligned} p_{mij} &\in [0, \mathrm{P}_{\max}] \;\; \forall\, (m,i,j) \in I\,, \\ x_{mij} &\in \{0,1\} \;\; \forall\, (m,i,j) \in I\,. \end{aligned} \tag{B.1}$$

For any assignment of these variables to form a feasible transmission schedule, we first need to ensure some simple, combinatorial conditions. Any station might only send *or* receive at the same time:

$$\sum_{j \in S} \sum_{m \in M} (x_{mij} + x_{mji}) \;\leq\; 1 \qquad \forall\, i \in S\,. \tag{B.2}$$

Furthermore, the semantics and relation between x and p variables have to be guaranteed, that is,

$$\mathrm{P}_{\min} x_{mij} \;\leq\; p_{mij} \;\leq\; \mathrm{P}_{\max} x_{mij} \qquad \forall\, (m,i,j) \in I\,. \tag{B.3}$$

The most complex set of constraint enforces the interference conditions (SINR). Since these conditions must only be active for actually scheduled transmissions ($x = 1$), we add a "switching" term containing a bound SG_∞ such that the constraint is only sharp if the referring transmission is scheduled. Writing γ_{ij} for the channel attenuation of transmission power between two stations $i, j \in S$ we obtain

$$\gamma_{ij} p_{mij} + \xi_m \mathrm{SG}_\infty (1 - x_{mij}) \geq \xi_m \left(\nu + \sum_{\substack{(n,s,r) \\ \neq (m,i,j)}} \gamma_{si} p_{nsr} \right) \quad \forall\ (m,i,j) \in I\,. \tag{B.4}$$

3.2 Time–Expanded Mixed–Integer Model

When compiling a transmission schedule, we have to ensure that feasible transmission conditions prevail at all times. A simple way to do so is to copy the basic model for each time slot. We get variables x_{mijt} and p_{mijt}, where

$$t \in \{1, \ldots, T\}$$

denotes the time slot, and we understand $x_{mijt} = 1$ as "Station i begins to transmit a data unit to station j at time t, using modulation m." Transmission will then be finished at time $t + l_m$. The construction is depicted in Fig. 2(b).

We furthermore take T copies of the constraints and substitute the term

$$\sum_{t \leq s < t + l(m)} x_{mijs}$$

for all occurrences of x_{mij} in the basic equations (B.2)–(B.4) at each time slot t. This provides the necessary link between overlapping transmissions. The time–expanded model including the formulation of the respective objective functions is presented in full detail in [2]. However, computational experiments showed that the model dimensions make the problem intractable for any nontrivial problem instances.

3.3 Column–Based Model

We now present a more sophisticated general model that uses the basic model for the pricing subproblem in a column generation scheme. The results, however, are approximations of actual transmission schedules and have to be rounded. The structure of this model is shown in Fig. 2(c).

Combining Transmission Patterns

The model we use has two tiers. The first tier, formulated as a linear program (LP), is responsible for finding a good distribution of the available time to all possible *transmission patterns*. A transmission pattern is a feasible set of concurrent transmissions and corresponds to a feasible solution $(\boldsymbol{x}^*, \boldsymbol{p}^*)$ of our basic model (B). The data throughput that each Station $i \in S$ achieves per time unit for any given transmission pattern is summarized in a matrix Φ. The column corresponding to $(\boldsymbol{x}^*, \boldsymbol{p}^*)$ is constructed by letting

$$(\Phi)_{i,*} := \sum_{j \in S} \sum_{m \in M} (x^*_{mij} - x^*_{mji})/l_m \qquad \forall\, i \in S\,. \tag{1}$$

Due to the constraints (B.2), at most one of the variables x in the sum is different from zero. The result $(\Phi)_{i,*}$ measures how much data station i puts on the network or receives from another station (negative data throughput) per time unit under a specific transmission pattern.

We write the problem of finding a time weighting $\boldsymbol{\eta}$ of all possible transmission patterns as a linear program. The matrix Φ is used to calculate the net data throughput that any station achieves. We sketch the model for the second fairness model (total user fairness), a full explanation and the model for the alternate fairness concept can be found in [2]. After adding an auxiliary variable $\bar{\alpha}$ to count the minimum data throughput our optimization problem reads

$$\begin{aligned} \max \quad & \bar{\alpha} \\ \text{s.t.} \quad & \mathbf{1}^T \boldsymbol{\eta} \leq T \\ & \Phi \boldsymbol{\eta} \geq \bar{\alpha} \\ & \boldsymbol{\eta} \geq 0 \end{aligned} \tag{LP}$$

Since we can only find a fractional time weighting $\boldsymbol{\eta}$ within this model, the resulting transmission schedule does not take into account that data units cannot be split in reality. A simple rounding-off procedure to cast the result to feasible transmission schedules with tolerable rounding loss is described in[2].

The Pricing Problem

The matrix Φ in the model (LP) has as many columns as there are feasible assignment of $\boldsymbol{x}$ variables in (B). These are too many to be considered explicitly, we therefore use a *column generation* approach. We start with a reduced set of columns, solve (LP), and calculate the dual variables $(y^*, \boldsymbol{w}^*)$ (where y corresponds to the time constraint and $\boldsymbol{w}$ to the constraints $\Phi\eta \geq \bar{\alpha}$). We use these dual variables to determine whether there are any columns that are not considered yet but could improve the objective of (LP). This problem is commonly called the *pricing problem* within the framework of column generation. This pricing problem, that is, the search for helpful transmission patterns, constitutes the second tier of our model.

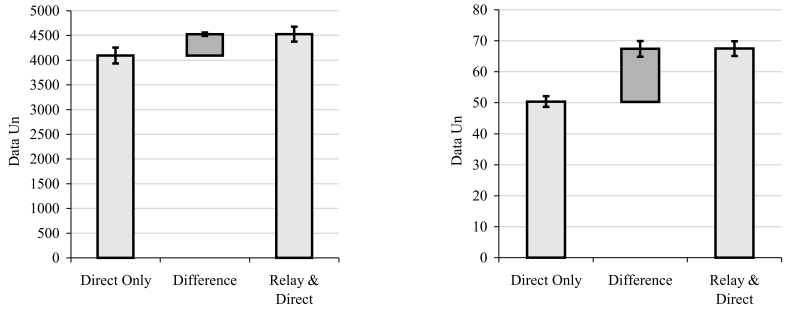
By writing down the dual problem of (LP) we see that $(y^*, \boldsymbol{w}^*)$ is feasible – which in turn means that the set of considered columns is sufficient – if and only if

$$\max_{\substack{\boldsymbol{x} \text{ feasible in (B)} \\ \text{with suitable } \boldsymbol{p}}} \sum_{(m,i,j)} \left(\frac{w_i^* - w_j^*}{l_m} \right) x_{mij} \leq y^* \,. \tag{PR}$$

We can thus solve the pricing problem by adding the left hand side of (PR) as the objective function to (B) and comparing the optimal value to y^*. If the inequality holds true, we are done. Otherwise we can use the vector $\boldsymbol{x}$ to construct a new column for (LP) as in (1).

4 Results

We implemented a column generation algorithm based on the model described in the previous section that computed almost optimal relaying schedules for special network configurations within a few hours on a standard PC (with an approximation of a few percent). The details can be found in [2]. The main computational work was to solve the pricing problem (PR) at late stages.



(a) Time share fairness: throughput gains of 10%

(b) Total user fairness: throughput gains of 30%

Fig. 3. Computational results

For selected classes of networks this implementation was used to quantify the average throughput gains by relaying. This was carried out by statistically analyzing the results for many "snapshots" of concrete configurations of mobiles hosts. For the first fairness model, fair time share, the result can be seen in Fig. 3(a); the average throughput gain of about 10% is rather small. This is because this fairness model favors mobiles that are close to the base station, and these mobiles are unlikely to use relaying. In the total user

fairness model, on the other hand, the gain amounts to about 30%. (All error bars represent 5% confidence intervals.)

Conclusions

We have presented mathematical programming models to solve a complex routing and scheduling problem for wireless cellular networks. These models were used to exemplarily show that the use of relaying can significantly improve data throughput in wireless LANs. This result represents a strong incentive to further develop the relaying approach on the technical side. Further work could, among others, focus on distributed algorithms or investigation of the influence of the fairness concept.

Acknowledgments

This work is based on a master's thesis written under the supervision of Dr. Holger Karl (Technical University Berlin, Telecommunications Network Group) and Dr. Andreas Eisenblätter and Dr. Arie Koster (Zuse Institute Berlin), to all of whom I owe thanks for extensive support. I also thank Prof. Dr. Martin Grötschel and Prof. Dr.Ing. Adam Wolisz.

References

1. P. Björklund, P. Värbrand, and D. Yuan. Resource optimization of spatial TDMA in ad hoc radio networks: A column generation approach. In *Proc. IEEE INFOCOM*, San Francisco, CA, March 2003.
2. Hans-Florian Geerdes. Capacity improvements in TDMA-based cellular networks by relaying and flexible transmission scheduling. Master's thesis, TU Berlin, 2002. Available online at `http://www.zib.de/groetschel/students/thesis_relaying.ps.zip`.
3. H. Karl and S. Mengesha. Analysing capacity improvements in wireless networks by relaying. In *Proc. of IEEE Intl. Conf. on Wireless LANs and Home Networks*, pages 339–348, Singapore, December 2001.
4. J. Khun-Jush, G. Malmgren, P. Schramm, and J. Torsner. HIPERLAN type 2 for broadband wireless communication. *Ericsson Review*, 2:108–119, 2000.
5. S. Kim, Z. Rosberg, and J. Zander. Combined power control and tranmission rate selection in cellular networks. In *Proc. of the 49th IEEE Vehicular Technology Conference Fall*, pages 1653–1657, Amsterdam, The Netherlands, September 1999.
6. Stavros Toumpis and Andrea Goldsmith. Capacity regions for wireless ad hoc networks. submitted, 2002.